Sample Complexity Bounds for RNNs with Application to Combinatorial Graph Problems

Nil-Jana Akpinar,[1] Bernhard Kratzwald,[2] Stefan Feuerriegel[2]
[1] Carnegie Mellon University, [2] ETH Zurich

Motivation

- Learning to predict solutions to real-valued combinatorial graph problems promises efficient approximations.
- Recurrent Neural Networks (RNNs) provide model-agnostic heuristics that can process graphs of variable sizes.
- Theoretical understanding of the number of samples required to learn real-valued RNNs, and thus their ability to scale for an exponentially increasing number of graphs, is missing.

Upper bound for RNN sample complexity

Theorem 1 (Sample complexity for single-layer RNNs)

A recurrent neural network with (1) a single recurrent layer of width a, (2) rectified linear units, (3) input of maximal length b, and (4) one real-valued output unit is learnable with sample complexity $M_L(\varepsilon, \delta)$ that is bounded by

$$M_L(\varepsilon,\delta) \leq \frac{128}{\varepsilon^2}\left[\ln\left(\frac{16}{\delta}\right) + \ln\left(\frac{34}{\varepsilon}\right) \times 4(a^2+3a+3)(2b(2a^2+4a)+4a+10+\log_2(8e))\right],$$

where ε is the population prediction error and $1-\delta$ the confidence.

Key implications:

- A population prediction error of ε can be obtained with at most $\tilde{\mathcal{O}}\left(a^4 b/\varepsilon^2\right)$ samples, i. e. linear in the maximum input length and polynomial in the number of recurrent units.
- Extension to d layers is straightforward and results in a sample complexity bound of $\tilde{\mathcal{O}}(d^2 a_{\max}^4 b/\varepsilon^2)$.
- Proofs are based novel bounds on RNN pseudo-dimension that make use of network size given by the number and type of parameters.

Application to combinatorial graph problems

- We aim at learning an approximation to a graph problem (such as ECCN) $\eta : \mathcal{G}_{\leq n} \to \mathbb{R}$, where $\mathcal{G}_{\leq n}$ is the set of graphs with up to n vertices.
- Each graph x is represented by its adjacency matrix mapped onto a single vector, i.e. $x = (x_1, \ldots, x_{m^2})$ $(m \leq n)$.

Theorem 2

Under some norm restrictions on parameters, a single hidden layer RNN with n units operating on $\mathcal{G}_{\leq n}$ can be learned with sample complexity $M_L(\varepsilon, \delta)$ bounded by

$$M_L(\varepsilon,\delta) \leq \frac{128}{\varepsilon^2}\left[\ln\left(\frac{16}{\delta}\right) + \ln\left(\frac{34}{\varepsilon}\right) \times 4(n^2+4n+3)(4n^4+8n^3+4n+10+\log_2(8e))\right],$$

where ε is the population prediction error and $1-\delta$ the confidence.

Key implications:

- A population prediction error of ε for our size-adaptive model can be reached with at most $\tilde{\mathcal{O}}(n^6/\varepsilon^2)$ samples, i. e. polynomial in graph size.
- For d layers, this only increases to $\tilde{\mathcal{O}}(d^2 n^6/\varepsilon^2)$ which suggest a favorable approximation performance is possible with a polynomial number of samples.
- This is a worst-case analysis independent of the problem distribution P and thus applicable to a wide range of graph problems.

The edge click cover number

- The Edge Click Cover Number (ECCN) of a graph is the minimum number of cliques, i. e. fully connected sub-graphs, required to cover all edges.
- NP-hard and thus computationally challenging.
- Relevant to a variety of practical applications incl. compiler optimization, protein interaction networks, etc.
- Known heuristics are scarce, with the Kellerman heuristic being the notable exception.

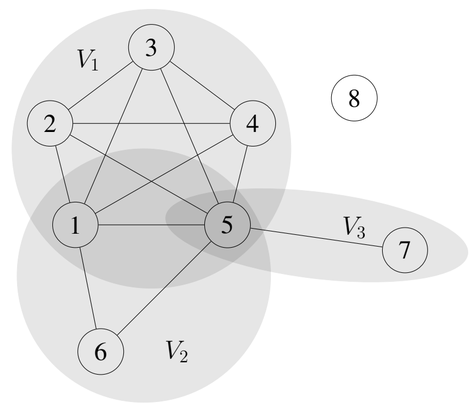

Figure 1: Graph with cliques $V_1 = \{1,2,3,4,5\}$, $V_2 = \{1,5,6\}$ and $V_3 = \{5,7\}$ forming a minimal edge clique cover.

Performance across training sample size

- We numerically evaluate how the number of training samples affects the out-of-sample performance in predicting the ECCN.
- With sufficient training data, neural learning over graphs is on par with the Kellerman heuristic and often even outperforms it.
- A reasonable performance can often be achieved with as little as 4,000 training samples.

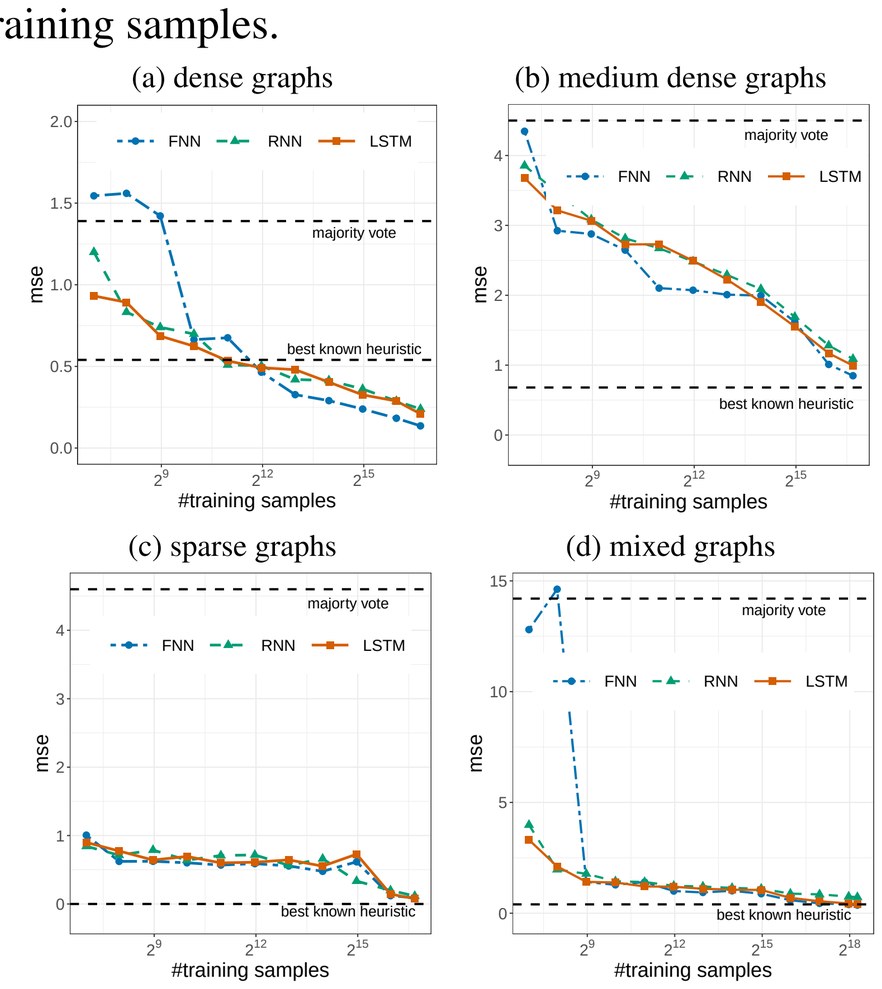

Figure 2: Out-of-sample performance in predicting ECCN. Mean squared error (mse) is reported as a function of the number of training samples (log-scale). With sufficient training samples, RNN predictions can outperform a naïve baseline and, in some cases, even the state-of-the-art heuristic.

Summary

- This is the first work that upper bounds the sample complexity for learning real-valued RNNs.
- Given a single-layer RNN with a rectified linear units and input of at most length b, we show that a population prediction error of ε can be realized with at most $\tilde{\mathcal{O}}(a^4 b/\varepsilon^2)$ samples. For d layers, $\tilde{\mathcal{O}}(d^2 a_{\max} b/\varepsilon^2)$ samples.
- A size-adaptive RNN fed with graphs of at most n vertices can be learned in $\tilde{\mathcal{O}}(n^6/\varepsilon^2)$, i. e., with only a polynomial number of samples.
- We demonstrate the effectiveness of RNNs on the NP-hard edge clique cover number problem, which even outperforms state-of-the-art heuristics.